# Core gene identification using gene expression

Claringbould, Annique

*DOI:*
10.33612/diss.145227875

# Appendices

## Summary

While humans share most of their genetic code with one another, changes in DNA can have a large impact on an individual's appearance, behaviour or risk of disease. In some cases, even small changes in one part of the DNA can lead to rare diseases. On the other hand, common genetic variants, such as SNPs, often only exert very small effects on the development of a disease or trait. Individually, these common variants do not lead to a disease, but their combined impact contributes to the heritable component of complex diseases. In the last 15 years, GWAS have identified thousands of SNPs that are associated with complex traits, like height, schizophrenia or arthritis.

However, knowing many of these genetic risk factors does not equate to understanding the molecular mechanisms that eventually lead to disease. One way to dive into the molecular changes that result from genetic variation is to look at changes in gene expression levels or DNA methylation. While the same genetic code is present in each cell, genes are only expressed when and where they are required (e.g. in particular cells or under particular circumstances). Genes are thus differentially expressed in various tissues, between individuals and upon stimulation. Over the last years it has become clear that many disease-associated genetic variants can also affect the expression of nearby (cis-eQTL) or distal (trans-eQTL) genes. Such a change in gene expression can lead to a slightly altered level of the protein the gene encodes, which in turn can be the start of dysregulation in the system that can eventually develop into a disease.

This thesis describes how gene expression patterns can be used to prioritise and characterise trait-relevant genes, leading to a better understanding of disease and potential drug targets. In the introduction (chapter 1), I describe the role of genetics in common complex traits and how genetic variations can be linked to gene expression.

In chapter 2, we evaluate the impact of genetic variation on methylation, gene expression, protein levels and various complex traits. After comparing the effect sizes of previously reported associations between SNPs and these data layers, we conclude that genetic variation can have a large effect on nearby molecular traits (cis-eQTLs and cis-meQTLs), while the effect on distal molecular traits (trans-QTLs) is smaller, and the impact on complex diseases is smaller still. This insight leads us to hypothesise that the distal effects are closer to disease on a scale of trait-complexity and might therefore prove more insightful than local effects.

Chapters 3 and 4 describe how to directly identify associations between gene expression or methylation on one hand and complex traits on the other. We use data from the BIOS Consortium, which has information on transcription, methylation and phenotypes of

thousands of individuals to compare the best analysis practices (chapter 3) and to zoom into aging specifically (chapter 4). These studies show that correction for blood cell composition (i.e. the quantity of each cell type within blood) is essential to get robust results.

In chapter 5, we aim to find the genes that play a causal role in determining the level of low-density lipoprotein levels in blood. Mendelian Randomisation is a statistical technique to find causal links, but it suffers from confounding when applied to gene expression levels. We therefore develop a new methodology, MR-link, that circumvents some of these issues and prioritise genes that are well-known to function in lipid metabolism.

Chapter 6 covers a large-scale study of how genetic variation affects gene expression levels in blood of 31,684 individuals. We find that SNPs associated to gene expression levels of nearby genes (cis-eQTLs) are very prevalent: 88% of all genes expressed in blood are regulated in this way. The distal trans-eQTLs usually have smaller effects, because they are indirect. Interestingly, we observe that variants associated to one disease often converge onto trans-eQTL genes relevant to that disease. We also calculate polygenic risk scores for 1,263 traits and identify associations between the combined risk and gene expression. These observations can be the starting point for further investigation into the molecular mechanisms of these diseases and of gene expression regulation in general. All results that are generated in this study are freely available on eqtlgen.org.

In chapter 7, we integrate publicly available gene co-expression data and GWAS summary statistics. Using this method, we are able to prioritise genes that are co-regulated with genes in or near GWAS loci. The rationale of this method is that thousands of disease-SNPs must work in concert to lead to a diseased phenotype, and this may well be captured by such co-expression patterns with core genes. The prioritised genes are shown to be more evolutionarily conserved and harbour fewer known mutations, indicating that they are indeed essential.

The last chapter describes the recently proposed omnigenic model, which states that genetic heritability of a disease can be spread across many (or even all) genes that are expressed in the relevant tissue, but that the disease mechanism is driven by a much smaller set of core genes. In this chapter, I describe how the work in this thesis fits within the framework of the omnigenic model.

The main take-home messages of this thesis are:

**1.** One method is rarely sufficient to prioritise locally causal genes, let alone core genes.
**2.** Cell composition likely influences most bulk gene expression studies, but certainly those that investigate gene expression in blood.
**3.** We need to continue to collaborate on large-scale (non-European) GWAS and tissue-specific datasets if we wish to use gene expression for identification of core genes.
**4.** Local gene regulation may only be trait-relevant if it is context-specific.
**5.** When used in the right context, Mendelian Randomisation can prioritise local causal genes.
**6.** While the omnigenic model probably simplifies the complexity of biology, it is a useful guideline to think beyond GWAS loci for gene prioritisation.
**7.** Core genes or their immediate neighbours will likely be useful as drug targets.

## Samenvatting

Iedereen op de wereld heeft bijna dezelfde genetische code. Toch kunnen de kleine verschillen in het DNA een groot effect hebben op iemands uiterlijk, gedrag of risico op een ziekte. In sommige gevallen kan zo'n klein verschil zelfs tot een zeldzame ziekte leiden. Aan de andere kant hebben veelvoorkomende genetische varianten, zoals *single nucleotide polymorphisms* (SNP's), vaak maar een beperkte invloed op het ontwikkelen van een ziekte of eigenschap. Elk van deze varianten op zich zal dus geen ziekte veroorzaken, maar gezamenlijk dragen ze wel bij aan de erfelijke component van multifactoriële ziekten. In de afgelopen 15 jaar hebben genoomwijde studies duizenden SNP's met allerlei eigenschappen en ziekten, zoals lengte, schizofrenie of artritis, kunnen associëren.

Kennis van deze genetische risicofactoren staat echter niet gelijk aan het begrijpen van de onderliggende moleculaire mechanismes die uiteindelijk tot zo'n ziekte leiden. Om het resultaat van die genetische variatie op moleculair niveau te onderzoeken, is het nuttig om de veranderingen in genexpressie en methylatie patronen te bestuderen. Elke cel bevat precies hetzelfde DNA, maar genen worden alleen tot expressie gebracht als ze op dat moment in de betreffende cel nodig zijn. Verschillende weefsels, individuen en omstandigheden hebben dus verschillende patronen van genexpressie. Het is door onderzoek van de afgelopen jaren duidelijk geworden dat veel van de eerdergenoemde genetische varianten die enigszins aan een ziekte bijdragen, ook een effect kunnen hebben op de expressieniveaus van dichtbijgelegen genen (*cis*-eQTLs) en van genen die ergens anders op het genoom liggen (*trans*-eQTLs). Zulke verschillen in genexpressie kunnen ervoor zorgen dat er net een beetje meer van het gecodeerde eiwit geproduceerd wordt. Die kleine ontregeling kan vervolgens de aanleiding zijn voor een reeks ontwikkelingen die uiteindelijk leiden tot het ontstaan van een ziekte.

In dit proefschrift beschrijf ik hoe die patronen van genexpressie gebruikt kunnen worden om ziekteverwekkende genen te prioriteren en te karakteriseren. Deze inzichten kunnen gebruikt worden om ziekten beter te begrijpen en medicijnen te ontwikkelen. In de introductie (hoofdstuk 1), beschrijf ik de rol van genetica in veelvoorkomende multifactoriële ziekten en hoe zulke genetische variatie aan genexpressie gelinkt kan worden.

In hoofdstuk 2 evalueren we het effect van genetische variatie op methylatie, genexpressie, eiwit niveaus en multifactoriële ziekten. Als we die analyses met elkaar vergelijken, kunnen we concluderen dat SNP's gemiddeld de grootste effecten hebben op genexpressie en methylatie van nabijgelegen genen (*cis*-eQTLs en *cis*-meQTLs), kleinere effecten op distale moleculaire kenmerken (*trans*-QTLs) en de kleinste effecten op ziekten. Door deze observaties veronderstellen we dat de distale effecten qua complexiteit dichter bij ziekten liggen dan de lokale effecten en wellicht daarom meer inzicht kunnen geven in de ontwikkeling van zo'n ziekte.

In hoofdstukken 3 en 4 beschrijven we de procedures voor het associëren van genexpressie of methylatie niveaus aan de ene kant met multifactoriële ziekten aan de andere kant. We gebruiken daarvoor data van het BIOS Consortium, dat voor elk van de duizenden deelnemers informatie bevat over transcriptie, methylatie en verschillende fenotypes. In hoofdstuk 3 vergelijken we analysemogelijkheden om de meest robuuste associatiemethode te vinden en in hoofdstuk 4 richten we ons specifiek op associaties tussen genexpressie en veroudering. Correctie voor bloedcelcompositie (m.a.w. de hoeveelheid van elk type cel in het bloed) blijkt essentieel voor het vinden van robuuste associaties.

Het doel van hoofdstuk 5 is om genen te vinden die een causale rol spelen bij de totstandkoming van het *low density* lipoproteïne niveau in het bloed. We gebruiken daarvoor *Mendelian Randomisation*, een statistische techniek die gebruikt wordt om causale verbanden te vinden tussen de blootstelling aan een risicofactor en de uitkomst. Als deze techniek toegepast wordt op genexpressie, leiden verstorende variabelen soms tot onbetrouwbare resultaten. Daarom hebben we een nieuwe methodologie ontwikkeld, MR-link genaamd, die deze problemen omzeilt. MR-link prioriteert voor het *low density* lipoproteïne niveau een aantal genen dat bekend staat om zijn rol in het lipide metabolisme.

In hoofdstuk 6 beschrijven we een grootschalige studie naar de effecten van genetische variatie op genexpressie niveaus in het bloed van 31,684 individuen. We vinden daarbij dat SNP's die de genexpressie van nabijgelegen genen beïnvloeden (*cis*-eQTLs) zeer frequent voorkomen: 88% van alle genen die in bloed tot expressie komen, worden op deze manier gereguleerd. De distale *trans*-eQTLs hebben meestal kleinere effecten, omdat de SNP een indirect effect heeft op het gen. We zien dat specifieke ziekte-SNP's vaak convergeren op *trans*-eQTL genen die voor die ziekte relevant zijn. We hebben eveneens polygene risicoscores uitgerekend voor 1,263 ziektes en eigenschappen en die geassocieerd met genexpressie. De genen die we met deze methodes prioriteren kunnen een uitgangspunt zijn voor verder onderzoek naar de precieze moleculaire toedracht van ziekten. Alle resultaten van deze studie zijn online beschikbaar op eqtlgen.org.

In hoofdstuk 7 staat de integratie van publiekelijk beschikbare gen co-expressie data met resultaten van genoomwijde associatie studies centraal. Met deze methode prioriteren we genen die gezamenlijk gereguleerd worden met genen die in (de buurt van) GWAS loci liggen. De gedachte hierachter is dat de duizenden bekende ziekte-SNP's op een bepaalde manier moeten samenwerken om tot een ziekte te leiden. Het zou goed kunnen dat GWAS genen co-regulatie vertonen met de *core* genen die er voor de ziekte echt toe doen. Aangezien de geprioriteerde genen evolutionair geconserveerd zijn en ze in de populatie minder vaak mutaties bevatten, veronderstellen we dat ze van essentieel belang zijn voor het organisme.

In het laatste hoofdstuk beschrijf ik het recent geïntroduceerde *omnigenic* model. Dit model postuleert dat de erfelijkheid van een ziekte weliswaar verspreid kan zijn over veel (of zelfs alle) genen die in het, voor die ziekte, relevante weefsel tot expressie komen, maar dat het werkelijke ziektemechanisme gedreven wordt door een veel kleiner aantal *core* genen. In hoofdstuk 8 beschrijf ik hoe de studies in dit proefschrift in het kader van het *omnigenic* model passen.

De belangrijkste conclusies van dit proefschrift zijn:

**1.** Het gebruik van slechts één methode is nagenoeg nooit voldoende om lokaal causale genen te prioriteren, laat staan *core* genen.
**2.** Celcompositie beïnvloedt waarschijnlijk veel van de genexpressie studies in verschillende weefsels, maar zeker alle onderzoeken naar genexpressie in bloed.
**3.** Als we genexpressie willen gebruiken om *core* genen aan te wijzen, moeten we (blijven) samenwerken om grootschalige (niet-Europese) GWAS en weefselspecifieke datasets te creëren.
**4.** Lokale regulatie van genen is mogelijk alleen relevant voor het begrijpen van ziekten als het in de juiste context bekeken wordt.
**5.** *Mendelian randomisation* kan, onder de juiste omstandigheden, lokaal de causale genen prioriteren.
**6.** Het *omnigenic* model simplificeert de complexiteit van biologie, maar is ondanks dat een nuttige leidraad om verder te kijken dan GWAS loci voor gen prioritering.
**7.** Het is aannemelijk dat *core* genen en hun naburige genen bruikbaar zullen worden als *drug targets.*

## Acknowledgments

It is quite strange to have just my name on the cover of this thesis. Science is a collaborative effort and it is, in fact, one of things I enjoy most about working in this field. The lists of wonderful co-authors on the chapters included in this thesis show that the work presented here is not something that I have accomplished on my own. On top of that, I am also indebted to a great number of people who are not directly involved with the scientific studies, but rather with making my years in Groningen such great ones. I would like to thank you all here.

### Department of Genetics

Lude, you have been such an inspiring and motivating supervisor. You helped me to see the big picture when I was struggling to understand statistical methods or fiddling with analysis details. I fondly remember many meetings where I would come in stuck on some problem and come out not only with a solution, but with lots of exciting new project directions. I am also grateful that you trusted me to lead, together with Urmo, the eQTLGen Consortium study. Because of that, I now understand what it means to pull off a large-scale collaborative multi-year research project, and while it has been frustrating at times, I also believe that we have set up a resource that will serve the scientific community well. So thank you for all you have taught me and I hope we get to brainstorm again every once in a while.

Cisca, thank you very much for your guidance during my PhD. I never understood how you managed, but despite your busy schedule (even more so since you've become the Rector), you always had time for a quick update at the coffee machine or a more in-depth discussion about science, career perspectives or life in general. On one of these talks you told me that I would 'land on my feet' with such confidence that I believed it myself as well. I really enjoyed and learned a lot from the PhD lunches over the years. You lift people up :)

I also want to thank my group members. It makes me happy to feel that we have each other's backs, both in science and outside of it. Urmo, our years-long collaboration has been (and still is) such a pleasant and effective one, even across country borders. You are meticulous, super R-savvy and you know when to stand your ground. I've learned a lot from you! Harm-Jan, you are exactly what I, and our group, needed when you came back to Groningen for your post-doc. You ask the right questions, you are a great sparring partner and, importantly, you know when to listen. Never sell yourself short. Patrick, thank you for teaching me how to take on scientific criticism. I will never forget a Monday lunch meeting where I'd received so many comments, also from you, that I could not finish my presentation within the timeslot. And then, afterwards, you came by to tell me that these comments meant that everyone wanted to help improve my work and that I should by no means take them as personal attacks. So then I didn't :)! Niek, from day 1 you were my PhD-buddy. We

started in Lude's group around the same time, went to the PhD introduction event together and wrote our first article at the department together. Your happy and content attitude is completely infectious and I don't think I've ever met anyone who is so sincerely happy for others' successes. I am so glad to call you my friend and am really looking forward to having you by my side as my paranymph. Dylan, you know how to bring people together and make everyone feel included, please don't ever close that big heart of yours. Thanks very much also to the other former and current FrankeSwertzLab members Dasha, Juha, Marc Jan, Sipko, Freerk, Monique, Harm, Roy, Robert, Damiano, Pauline, Floranne, Irene, Joeri, Tyler, Alex and Shuang for making team science so rewarding. Thanks to the students I've supervised: Vera, Gerjanne and Maartje and those from the Data Science for Life Sciences master programme. Explaining the concepts of genetics to you has made me a better scientist. Hanna and Anil, you are both the type of scientist I look up to: very kind and very smart. Thanks for taking me along in the world of brains & genetics.

Over the years I've seen many PhD students come and go to the department of genetics and you all have made it a great place to work. Thank you for the PhD lunches, Game of Thrones watch-parties and now-I'm-a-true-nerd LAN Factorio pizza nights. Kieu, your smile and positive attitude always light up my day and our dancehall classes were so much fun! Kai, Olivier, Dylan, Niek, Sipko and Monique: thanks for the many fun board game and DnD (yep..) nights together. Arnau, Raúl, Vicky, Maria, Esteban, Aaron, Werna and Adriaan: I really enjoyed working and not-working with you. To the (former) PIs in the department, Iris, Sasha, Morris, Jingyuan, Yang, Serena, Vinod and Sebo, your ideas and talent are what drives the department forward and I feel very lucky to have met and worked with you. Thanks to Nine for heading the department with a work-hard-play-hard motto. From the GCC I would like to thank Pieter for patiently answering at least a thousand of my questions, Roan, Gerben, Marieke and Sido for helping out at various stages of the projects and Dennis for his unwavering confidence. I hope to see you again in De Koffer! Kate, your editing has been instrumental for getting the message of this thesis across, thank you. Your open attitude during the PhD lunch made it my favourite one and I hope you continue to follow your writing dream. Janneke, bij jou binnenkomen zorgde ervoor dat ik me op de afdeling thuis voelde.

**Collaborations**

During my PhD, I have been fortunate to also work with many people outside of our own department. Jian Yang, Peter Visscher, Naomi Wray and the whole team at PCTG: thank you for hosting me in Brisbane for four months. I learned so much about statistical genetics from you. Peter and Naomi, thank you for welcoming me into the group and your home from the very first week, it made all the difference. I would like to sincerely thank De Drie Lichten for their travel grant that made it possible for me to spend that time in Australia. Many thanks also to Bogdan Pasaniuc and the great team at UCLA for inviting me to join the fantastic computation genomics summer institute in 2019 and for hosting me in your group thereafter. Your approach to complex trait genetics is very exciting and I am very grateful that you gave me the opportunity to work alongside you for a while.

Almost every chapter in this thesis relies on data from research participants. To all those individuals that participated in any of the eQTLGen cohorts, the BIOS cohorts and specifically the Groningen-based LifeLines cohort: thank you. This work could not have been done without the selfless contribution that you made in the name of science.

**Friends and family**

I think it was the beginning of 2016 when I joined the PhD introduction event with some 30 other fresh PhD students from the University of Groningen. These two days led to the formation of what we have come to call Team Groningen: a group of friends that made their way through their early careers around the same time I did. From camping trips (sssshhh!) to celebrating Sinterklaas and from shared writing struggles to late nights in the city centre, you have been a fantastic group of friends that made me feel at home in Groningen. Thank you so much to Hannah, Joana, Paolo, Désirée, Steven, Adele, Tejas, Lucile, Filippo, Alberto, Sharon, Vincent, Niek and Anne-Grete!

Dear Anne-Grete, I feel like we were friends almost from the moment I met you. You are so incredibly smart and dedicated to your work that at your current rate your thesis will include about 30 chapters. But more importantly, you are warm and inviting and such a good friend. I cherish all the dinners, glasses of wine and drilling-into-your-house sessions we've had and I am excited and thankful that you are my paranymph.

Lieve Sophie, ik vind het zo bijzonder dat we na al die jaren nog steeds besties zijn! Ik ben dan ook ontzettend blij dat je deel bent van dit proefschrift. Dankjewel voor het ontwerpen van deze prachtige, vrolijke en artistieke omslag. Ik zal altijd een beetje aan jou en je creativiteit denken als ik naar dit boekje kijk en dat maakt me nu al blij.

Lieve papa en mama, jullie vertrouwen, liefde en advies zijn van heel veel betekenis voor mijn leven in het algemeen en voor het behalen van deze graad in het bijzonder. Jullie hebben vaak geluisterd naar uitleg over mijn onderzoek en het proces van wetenschap, maar waren net zo goed begripvol als ik het er even niet over wilde hebben. Jullie ondersteunen me in elke keuze die ik maak en twijfelen nooit aan de goede afloop. Dank jullie wel voor het meelezen en meeleven. Lieve Marnix, *thanks* voor je muziektips (heel fijn tijdens werk), klimsessies en voor je *reassuring attitude*. Lieve oma, bedankt voor alle interesse in en trots op mijn werk.

Lieve Mischa, het leven is zo fijn met jou aan mijn zijde. Dankjewel dat je er altijd bent om met me te sparren of me te laten *venten*, dat je mijn werk en keuzes respecteert en dat je de kalmte uitstraalt die ik af en toe nodig heb. Ik kan niet wachten om samen meer van de wereld te ervaren.

## Curriculum vitae

Annique Juliëtte Claringbould was born on January 28th 1992 in 's-Hertogenbosch, the Netherlands. After graduating *cum laude* from the Stedelijk Gymnasium 's-Hertogenbosch in 2010, she went on to study Liberal Arts and Sciences at University College Utrecht. She majored in life sciences and cognitive neuroscience and completed a minor in philosophy. She graduated *cum laude* in 2013. Annique became interested in the field of genetics during her bachelor thesis internship at the University Medical Centre Utrecht, where she studied de novo mutations in rare disease patients in the lab of dr. Gijs van Haaften as well as the ethical benefits and disadvantages of sharing DNA data under supervision of dr. Annelien Bredenoord.

In 2014, she was accepted into the Human Molecular Genetics MSc programme at Imperial College London with a full scholarship. During her master internship in dr. Inga Prokopenko's group, she performed multivariate genome-wide association studies and discovered that she wanted to pursue statistical genetics. Annique was awarded the prizes for best course performance and best research project and graduated with distinction. She presented her MSc thesis research at the Quantitative Genomics conference in London, UK, the Functional Genomics and Systems Biology conference in Hinxton, UK, and the International Congress of Human Genetics conferences in Kyoto, Japan.

Annique started her PhD under supervision of prof. Lude Franke and prof. Cisca Wijmenga at the Department of Genetics of the University Medical Centre Groningen in the fall of 2015. She researched gene expression patterns and how they can be leveraged to find disease-causing genes, resulting in this thesis. In 2018, Annique spent 4 months in the lab of prof. Jian Yang at the University of Queensland, Australia, supported by a travel grant by De Drie Lichten. She was also hosted by dr. Bogdan Pasaniuc at the University of California at Los Angeles for a summer work visit in 2019. During her PhD, Annique presented her work at the European (2017) and American (2018) Society of Human Genetics conferences and was an invited speaker at the 2019 Quantitative Genetics and Genomics Gordon Research Conference.

Annique is currently working as a postdoctoral fellow at the European Molecular Biology Laboratory in Heidelberg, Germany, in the lab of dr. Judith Zaugg and co-hosted by dr. Lars Steinmetz.

## List of Publications

### First Author Publications

- Pellegrino Coppola, D.*, Claringbould, A.*, *et al.* (2020) 'Correction for both common and rare cell types in blood is important to identify genes that 1 correlate with age 2 3', *bioRxiv*, p. 2020.05.28.120600. doi: 10.1101/2020.05.28.120600.
- Võsa, U.*, Claringbould, A.*, *et al*. (2018) 'Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis', *bioRxiv*, p. 447367. doi: 10.1101/447367.
- Claringbould, A.*, de Klein, N.* and Franke, L. (2017) 'The genetic architecture of molecular traits', *Current Opinion in Systems Biology.* doi: 10.1016/j.coisb.2017.01.002.

### Second Author Publications

- Aguirre-Gamboa, R.*, de Klein, N.*, Tommaso, J.*, *et al.* (2020) 'Deconvolution of bulk blood eQTL effects into immune cell subpopulations', *BMC Bioinformatics*, 21(1), p. 243. doi: 10.1186/s12859-020-03576-5.
- Graaf, A. van der, et al. (2020) 'Mendelian randomization while jointly modeling cis genetics identifies causal relationships between gene expression and lipids', Nature Communications 2020 11:1, 11(1), pp. 1–12. doi: 10.1038/s41467-020-18716-x.
- Porcu, E. *et al.* (2020) 'The role of gene expression on human sexual dimorphism: too early to call', *bioRxiv*, p. 2020.04.15.042986. doi: 10.1101/2020.04.15.042986.
- van Rooij, J.*, Mandaviya, P.R.*, *et al.* (2019) 'Evaluation of commonly used analysis strategies for epigenome- A nd transcriptome-wide association studies through replication of large-scale population studies', *Genome Biology*, 20(1), p. 235. doi: 10.1186/s13059-019-1878-x.

## Co-author Publications

- Blokland, I. V. van *et al.* (2020) 'Using symptom-based case predictions to identify host genetic factors that contribute to COVID-19 susceptibility', *medRxiv*, p. 2020.08.21.20177246. doi: 10.1101/2020.08.21.20177246.
- Chen, J. *et al.* (2020) 'The Trans-Ancestral Genomic Architecture of Glycaemic Traits', *bioRxiv*. doi: 10.1101/2020.07.23.217646.
- Folkersen, L. et al. (2020) 'Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals', Nature Metabolism, 2(10), pp. 1135–1148. doi: 10.1038/s42255-020-00287-2.
- Graaf, A. van der *et al.* (2020) 'Systematic prioritization of candidate genes in disease loci identifies TRAFD1 as a master regulator of IFNγ signalling in celiac disease', *bioRxiv*, p. 2020.03.04.973487. doi: 10.1101/2020.03.04.973487.
- de Klein, N.*, van Dijk, F.*, et al. (2020) 'Imbalanced expression for predicted high-impact, autosomal-dominant variants in a cohort of 3,818 healthy samples', bioRxiv, p. 2020.09.19.300095. doi: 10.1101/2020.09.19.300095.
- Kurilshikov, A. *et al.* (2020) 'Genetics of human gut microbiome composition', *bioRxiv*, p. 2020.06.26.173724. doi: 10.1101/2020.06.26.173724.
- Mc Intyre, K. *et al.* (2020) 'The Lifelines COVID-19 Cohort: a questionnaire-based study to investigate COVID-19 infection and its health and societal impacts in a Dutch population-based cohort', *medRxiv*, p. 2020.06.19.20135426. doi: 10.1101/2020.06.19.20135426.
- Ouwens, K. G. *et al.* (2020) 'A characterization of cis- and trans-heritability of RNA-Seq-based gene expression', *European Journal of Human Genetics*, 28(2). doi: 10.1038/s41431-019-0511-5.
- Alves, A. C. *et al.* (2019) 'GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI', *Science Advances*, 5(9). doi: 10.1126/sciadv.aaw3095.
- Prokopenko, I. *et al.* (2019) 'Alzheimer's disease pathology explains association between dementia with Lewy bodies and APOE-ε4/TOMM40 long poly-T repeat allele variants', *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, 5. doi: 10.1016/j.trci.2019.08.005.
- Schlicht, K. *et al.* (2019) 'The metabolic network coherence of human transcriptomes is associated with genetic variation at the cadherin 18 locus', *Human Genetics*, pp. 1–14. doi: 10.1007/s00439-019-01994-x.
- Zeng, B. *et al.* (2019) 'Comprehensive Multiple eQTL Detection and Its Application to GWAS Interpretation.', *Genetics*, p. genetics.302091.2019. doi: 10.1534/genetics.119.302091.
- Luijk, R. *et al.* (2018) 'Genome-wide identification of directed gene networks using large-scale population genomics data', *Nature Communications*, 9(1), p. 3097. doi: 10.1038/s41467-018-05452-6.
- Yousefi, S. *et al.* (2018) 'A SNP panel for identification of DNA and RNA specimens', *BMC Genomics*, 19(1), p. 90. doi: 10.1186/s12864-018-4482-7.
- Zhernakova, D. V. *et al.* (2018) 'Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome', *Nature Genetics*, 50(11), pp. 1524–1532. doi: 10.1038/s41588-018-0224-7.

## eQTLGen Consortium publications

The eQTLGen Consortium has collaborated with many research groups and is included as a banner author on the following articles.


- Hammerschlag, A. R. *et al.* (2020) 'Refining Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Genetic Loci by Integrating Summary Data From Genome-wide Association, Gene Expression, and DNA Methylation Studies', *Biological Psychiatry*, 88(6), pp. 470–479. doi: 10.1016/j.biopsych.2020.05.002.
- Mills, M. C. *et al.* (2020) 'Identification of 370 loci for age at onset of sexual and reproductive behaviour, highlighting common aetiology with reproductive biology, externalizing behaviour and longevity', *bioRxiv*, p. 2020.05.06.081273. doi: 10.1101/2020.05.06.081273.
- Karlsson Linnér, R. *et al.* (2019) 'Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences', *Nature Genetics*, 51(2), pp. 245–257. doi: 10.1038/s41588-018-0309-3.
- Männik, K. *et al.* (2019) 'Leveraging biobank-scale rare and common variant analyses to identify ASPHD1 as the main driver of reproductive traits in the 16p11.2 locus', *bioRxiv*, p. 716415. doi: 10.1101/716415.
- Porcu, E. *et al.* (2019) 'Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-10936-0.
- Thompson, D. J. *et al.* (2019) 'Genetic predisposition to mosaic Y chromosome loss in blood', *Nature*. doi: 10.1038/s41586-019-1765-3.
- Timmers, P. R. *et al.* (2019) 'Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances', *eLife*, 8. doi: 10.7554/eLife.39856.
- Qi, T. *et al.* (2018) 'Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-04558-1.
- Wray, N. R. *et al.* (2018) 'Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression', *Nature Genetics*, 50(5), pp. 668–681. doi: 10.1038/s41588-018-0090-3.
- Xue, A. *et al.* (2018) 'Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes', *Nature Communications*, 9(1), p. 2941. doi: 10.1038/s41467-018-04951-w.
- Lepik, K. *et al.* (2017) 'C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis', *PLOS Computational Biology*. Edited by A. Rzhetsky, 13(9), p. e1005766. doi: 10.1371/journal.pcbi.1005766.


* equal contribution